Applied Survival Analysis
Lab 9: Time-dependent covariates

In today's lab, we are going to get familiar with time-dependent covariates. We will use the Stanford heart transplant dataset *(naïve.dta).* In this study 103 patients waiting for a heart transplant were followed for survival. Here is a description of the file:

```
describe
  obs:           103
 vars:            12                          19 Dec 2003 15:31
 size:         5,356 (99.5% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
patid           float  %9.0g                  Patient identifier
year            float  %9.0g                  Year of acceptance
age             float  %9.0g                  Age
fail            float  %9.0g                  Survival status (1=dead)
survtime        float  %9.0g                  Survival time
priorsurg       float  %9.0g                  Surgery
transplant      float  %9.0g                  Heart transplantation  status
                                                (1=yes)
waitime         float  %9.0g                  Waiting time for transplant
missallele      float  %9.0g
missantigen     float  %9.0g
misscore        float  %9.0g
time            float  %9.0g                  Survival time (correction for
                                                id=38)
-------------------------------------------------------------------------------
Sorted by:  patid
```

There is one record for each patient and the important variables in the framework of time-dependant analysis are **id, transplant, waitime, fail and time**.

A naive approach for estimating the effect of transplantation to the hazard ratio of having vs not having a new heart is to apply a Cox model to the above dataset by considering the transplantation status as a fixed binary covariate.

```
. stset time, f(fail)

     failure event:  fail != 0 & fail < .
obs. time interval:  (0, time]
 exit on or before:  failure

------------------------------------------------------------------------------
      103  total obs.
        0  exclusions
------------------------------------------------------------------------------
      103  obs. remaining, representing
       75  failures in single record/single failure data
  31948.1  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      1799


. stcox  transplant, nohr

         failure _d:  fail
   analysis time _t:  time

Iteration 0:   log likelihood = -298.31514
Iteration 1:   log likelihood =   -287.817
Iteration 2:   log likelihood = -285.44061
Iteration 3:   log likelihood = -285.44037
Refining estimates:
Iteration 0:   log likelihood = -285.44037

Cox regression -- Breslow method for ties
No. of subjects =         103                  Number of obs   =       103
No. of failures =          75
```

```
Time at risk    =      31948.1
                                                  LR chi2(1)      =     25.75
Log likelihood  =   -285.44037                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |  -1.318887   .2439952    -5.41   0.000    -1.797109   -.8406655
------------------------------------------------------------------------------
```

The estimated coefficient is associated with a hazard ratio of 0.267, which implies that patients with a new heart have about 4 times less hazard to die than those without a new heart. This analysis handles transplantation status as a fixed covariate and does not account for the fact that a very high hazard is likely to follow transplantation.
To perform an analysis involving time-updated transplant status we need to transform the dataset by creating multiple lines per subject.
To understand the structure of the dataset that we are going to produce, let's take a look at the original data.

```
. li patid  transplant waitime survtime time fail if patid==38 | patid==16 |
pat==12 | pat==80, clean

        patid   transp~t   waitime   survtime   time   fail
          12          0         .          8      8      1
          16          1        28        308    308      1
          38          1         5          5    5.1      1
          80          1        26        482    482      0
```

Patient 12 never received a new heart. He died 8 days after acceptance while still on the waiting list. Patient 80 did receive a new heart 26 days after acceptance and he survived until the end of follow-up. Patient 38 died the day of the heart transplantation, 5 days after acceptance. Such cases would be excluded from the statistical software, so we add a small fraction (0.1) to the survival time (time=5.1 instead of 5)

Our goal is to turn this into a dataset that contains the histories of each patient, that is we want records that appear as follows:

```
        patid   transp~t   waitime   time   fail
          12          0         .      8      1
          16          0        28     28      0
          16          1        28    308      1
          38          0         5      5      0
          38          1         5    5.1      1
```

This means that for patients who get a new heart we want to have 1 record for the follow-up period before the transplantation and 1 record for the period after. For this reason we apply the following commands:

```
expand 2 if transplant
(69 observations created)
sort patid
```

This will cause a dublication of the records for patients who had transplant=1. The new file contains 172 observations.

```
. li  patid  transplant  waitime time fail if pat==12 | pat==16 | pat==38 | pat==80,
clean
       patid   transp~t   waitime   time   fail
 17.      12          0         .      8      1
 23.      16          1        28    308      1
 24.      16          1        28    308      1
 60.      38          1         5    5.1      1
 61.      38          1         5    5.1      1
132.      80          1        26    482      0
133.      80          1        26    482      0
```

So we got 1 observation for patient 12 who had transplant=0 and 2 for the rest who had transplant=1. The problem is that patients with 2 records have identical values for all the variables, so **time, fail and transplant** (the time-dependent covariate in this example) do not reflect the transplantation and failure history. Within the records of each patient we replace this variables as follows:

```
by patid:replace time=waitime if _n==1 & transplant
by patid:replace fail=0 if _n==1 & trans
by patid:replace transplant=0 if _n==1 & transplant
```

Thus the above records are now:

```
       patid   transp~t   waitime   time   fail
 17.      12          0         .      8      1
 23.      16          0        28     28      0
 24.      16          1        28    308      1
 60.      38          0         5      5      0
 61.      38          1         5    5.1      1
132.      80          0        26     26      0
133.      80          1        26    482      0
```

Our data have now the desired form. Save this file with a new name e.g. *transplant.dta.* Let' s now stset the new dataset.

```
. stset  time, f(fail) id(patid)
                id:  patid
     failure event:  fail != 0 & fail < .
obs. time interval:  (time[_n-1], time]
 exit on or before:  failure

------------------------------------------------------------------------------
      172  total obs.
        0  exclusions
------------------------------------------------------------------------------
      172  obs. remaining, representing
      103  subjects
       75  failures in single failure-per-subject data
  31948.1  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      1799
```

The option id(patid) specifies the subject-id variable and indicates that observations with the same id refer to the same patient.

To estimate the effect of heart transplantation we apply the following Cox model.

```
. stcox  transplant, nohr
```

```
         failure _d:  fail
   analysis time _t:  time
                 id:  patid

Iteration 0:   log likelihood = -298.31514
Iteration 1:   log likelihood = -298.25194
Iteration 2:   log likelihood = -298.25193
Refining estimates:
Iteration 0:   log likelihood = -298.25193

Cox regression -- Breslow method for ties

No. of subjects =          103                     Number of obs   =       172
No. of failures =           75
Time at risk    =      31948.1
                                                   LR chi2(1)      =      0.13
Log likelihood  =   -298.25193                     Prob > chi2     =    0.7222

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |   .1057311   .2984149     0.35   0.723    -.4791513    .6906136
------------------------------------------------------------------------------
```

The estimated coefficient corresponds to a hazard ratio of 1.11 in favor of subjects who did not take a new heart, that is after heart transplantation the hazard for death increases by 11%.